

Machine Learning vs. Conventional Statistics in Predicting Medical Outcomes: A Comparative Performance Analysis Using a Viral Disease Dataset

✉ Fatih İkiz¹, ✉ Ahmet Ak²

¹University of Health Sciences Türkiye, Konya Beyhekim Training and Research Hospital, Clinic of Emergency Medicine, Konya, Türkiye

²Selçuk University Faculty of Medicine, Department of Emergency Medicine, Konya, Türkiye

Abstract

Aim: Accurate mortality prediction is fundamental for clinical resource allocation and personalized patient management across diverse medical conditions in patients admitted to emergency department. This study evaluates the predictive capacity of laboratory and clinical variables by conducting a direct comparison between conventional statistical models and advanced machine learning (ML) algorithms.

Materials and Methods: Mortality-associated variables were first analyzed using conventional approaches, including basic comparative tests, logistic regression, and ROC analyses. Subsequently, eleven ML algorithms were deployed to benchmark their performance against these conventional methods. The dataset was partitioned into a 75% training set and a 25% testing set. Models were evaluated based on sensitivity (recall), area under the curve (AUC), and overall accuracy. Features of importance were defined as ranks assigned by the most robust model to identify key clinical predictors.

Results: While conventional statistical methods achieved a maximum sensitivity of 81.2% (AUC=0.906), ML algorithms significantly outperformed conventional statistical methods. Among the eleven algorithms, BayesNet emerged as the superior model, with a sensitivity of 88.9% and an overall classification accuracy of 92.3%. The analysis demonstrates that ML techniques capture complex, non-linear interactions within clinical data that standard logistic regression may overlook.

Conclusion: ML algorithms offer a substantial improvement in predictive performance over conventional statistical methods for mortality prediction in patients with complex viral diseases admitted to the emergency department. These findings suggest that integrating ML into clinical decision-support systems can provide more precise risk stratification than traditional prognostic tools.

Keywords: Artificial intelligence, machine learning, mortality, logistic regression, BayesNet

Introduction

The complex, multifactorial nature of viral diseases such as coronavirus disease 2019 (COVID-19), involving intricate interactions between numerous clinical and laboratory variables and potentially exhibiting non-linear relationships, may inherently limit the predictive capacity of these conventional, linear-based statistical approaches. Statistical methods, although their analytical forms can be more readily controlled and analyzed, depend on assumptions for their feasibility. Limitations may arise during the evaluation of numerous independent variables and multifactorial assessments (1).

In recent years, the field of machine learning (ML) has witnessed remarkable advancements, offering a suite of sophisticated algorithms capable of discerning complex patterns and making accurate predictions from high-dimensional datasets. In the realm of medicine, ML techniques such as decision trees, random forests, gradient boosting machines, support vector machines, and artificial neural networks have demonstrated considerable promise in various diagnostic and prognostic applications (2). Their ability to capture complex relationships, automatically identify feature interactions, and handle large volumes of data presents a compelling alternative and potential improvement



Corresponding Author: Fatih İkiz MD, University of Health Sciences Türkiye, Konya Beyhekim Training and Research Hospital, Clinic of Emergency Medicine, Konya, Türkiye
E-mail: sultanmehmet01@hotmail.com **ORCID ID:** orcid.org/0000-0003-3710-2193

Cite this article as: İkiz F, Ak A. Machine learning vs. conventional statistics in predicting medical outcomes: a comparative performance analysis using a viral disease dataset. Eurasian J Emerg Med. 2026;25: 228-36.

Received: 02.01.2026

Accepted: 22.03.2026

Published: 30.03.2026



©Copyright 2026 The Author(s). The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House.
Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License

over conventional statistical methods, which still require further investigation. A comprehensive comparison focusing on the synergistic predictive value of coagulation parameters combined with standard clinical variables remains an area for further investigation.

Because it is a multifaceted and multifactorial disease and the effects of coagulation parameters on clinical outcomes have been demonstrated, the COVID-19 dataset was selected for use in this study. The dataset was deemed suitable for our comparison model for these reasons: it contains a sufficient sample size and well-defined clinical and laboratory outputs. To our knowledge, this is the first study to compare ML implementations with traditional statistical methods in this way for any viral disease, including COVID-19. The number of studies investigating the primary success rates of ML and statistical methods in medicine is limited. In this respect, we believe that our study will make a significant contribution to the literature and offer important insights to researchers regarding methodology selection.

Materials and Methods

Our study was conducted in the emergency department and focused on prognostic outcomes of patients, including their initial admission data. The study protocol was designed in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Selçuk University Rectorate Local Ethics Committee of a university hospital (approval no: 2025/218, date: 08.04.2025). The data between 19.03.2020 and 26.10.2020 were accessed retrospectively with permission. Therefore, following ethical approval, data were retrospectively accessed between 08.04.2025 and 08.05.2025 and were obtained from patients aged 18 years and older who were admitted to the university hospital's emergency department and diagnosed with COVID-19 between 19.03.2020 and 26.10.2020. The authors had access to medical records that could identify individual participants during data collection. Because the study is retrospective and the data were provided by the hospital anonymously (with no names and confidential info), no consent was obtained in this regard. The research was concluded on 01.06.2025.

Research Model

This retrospective case-control study was conducted using data from 620 patients diagnosed with COVID-19 during the pandemic period between 19.03.2020 and 26.10.2020. Our study utilized coagulation markers to investigate parameters associated with 30-day in-hospital mortality [D-dimer, fibrinogen, prothrombin time (PT), activated PT (aPTT), and platelet]. Clinical findings glasgow coma scale (GCS), comorbidities, computerized tomography (CT) findings, and vital parameters at first admission were also

assessed, which may be related to mortality. PCR positivity has been accepted as a criterion for COVID-19 diagnosis in our study. Ground-glass opacities, consolidations, and paving patterns are considered positive CT findings suggestive of COVID-19. The relevant findings were investigated using statistical methods and ML models, and the results were evaluated by a comprehensive comparison (conventional statistics vs. ML) (Figure 1).

Statistical Analysis

Statistical analyses were performed using SPSS 27.0 (IBM Inc., Chicago, IL, USA). The Kolmogorov-Smirnov test, histogram analyses, skewness and kurtosis, and Q-Q plots were used to evaluate the normality of continuous variables. Continuous variables were expressed as interquartile ranges [median (minimum-maximum)] or mean \pm standard deviation. Qualitative data are expressed as frequencies (N) and percentages (%). Relationships of continuous variables between two groups were examined using the Mann-Whitney U test or the independent-samples t-test. The effect profiles of variables on mortality were evaluated using univariate and multivariate logistic regression analyses, limited to 10 events per variable. Box-Tidwell and multicollinearity assumptions were assessed prior to multivariate logistic regression analysis. Relationships between qualitative variables were investigated using Pearson's chi-square test or Fisher's exact test. Throughout the study, the type I error rate was set at 5% ($\alpha=0.05$), and p-values less than 0.05 were considered statistically significant.

ML Methods

The survival and non-survival groups were separated to ensure homogeneous class distributions in the training and test sets. Subsequently, the sample was randomly divided into two independent groups: a 75% training set and a 25% test set. The data used in the test set were not used in the training set. To ensure that the models were trained and evaluated on representative samples, this split was performed using stratified sampling based on the binary outcome variable (survival/non-survival groups). This stratification ensured that the ratio of survival to non-survival classes was maintained in both the training and test sets. The test set was strictly held out and used only once for the final unbiased evaluation of the best model selected. Model development and initial comparison were conducted entirely on the 75% training set using 5-fold stratified cross-validation.

ML models and data mining were carried out and evaluated using WEKA software, version 3.8.6. ML models were run on a Windows 11 operating system using an Intel i7 CPU, 16 GB of RAM, and an NVIDIA GTX 1660 Ti 8 GB graphics card. The selection of ML algorithms was designed to represent a broad spectrum of computational approaches, including Bayesian

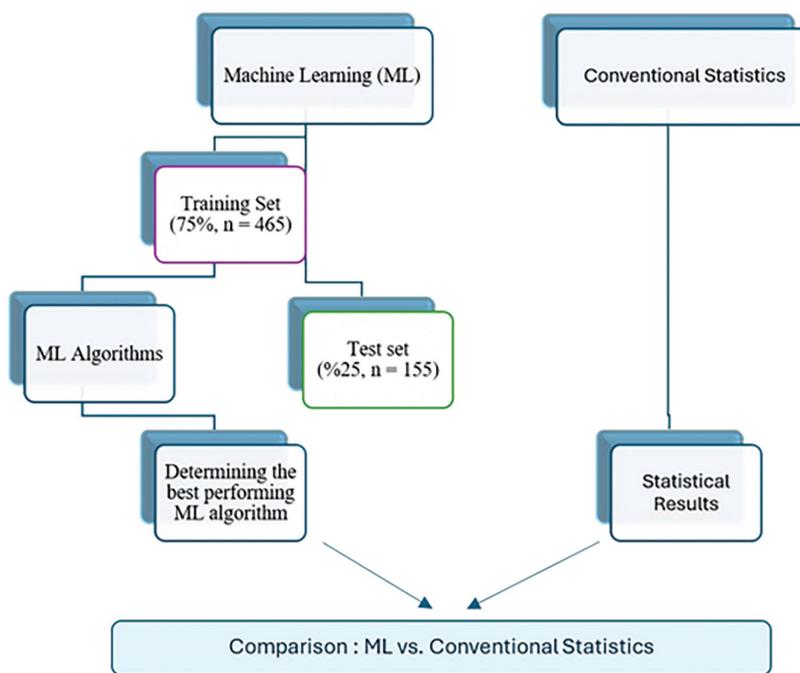


Figure 1. General flow chart of the study

ML: Machine learning

networks, decision trees, rule-based classifiers, and functional models. This diversity ensures that the comparison accounts for both linear and non-linear data structures. Furthermore, these specific algorithms were chosen based on their established performance in the medical prognostic literature and their availability within the WEKA 3.8.6 framework, thereby allowing standardized, reproducible benchmarking against conventional statistical methods. The 11 most popular ML models suitable for our study’s model and dataset were applied and reported. To evaluate the predictive capabilities of the ML algorithms, mean absolute error, root mean squared error, correlation coefficients, Matthews correlation coefficient, F1 score, and recall were reported. The most successful ML algorithm was selected based on achieving the highest area under the curve (AUC) and F1-score. The AUC was chosen as the primary metric to select the model with the greatest overall discriminatory power, and the F1-score-the harmonic mean of sensitivity (recall) and precision-was used as a secondary measure to confirm a clinically relevant balance between minimizing false negatives and false positive (FP). In cases where a clear trade-off existed between the two metrics, the algorithm with the higher AUC was prioritized. Metrics such as true positive (TP) rate, FP rate, precision, and AUC were assessed for clinical utility. Among the algorithms with acceptable overall accuracy (>90%) and F1 scores (>0.70), the algorithm with the highest AUC value-also taking into account

clinical adaptability and value in the medical field-was deemed the most suitable when evaluated from these perspectives. In this regard, it was concluded that the BayesNet algorithm was optimal. For the BayesNet, metrics such as the direction of use of the relevant attributes (variables) and feature importance in their prediction of the relevant class were reported as ranks. The structure of the final BayesNet was determined using the K2 algorithm with a batch size of 100, and the feature importance and direction were determined by analyzing the conditional probability tables (CPTs). The CPTs were estimated using the SimpleEstimator function (weka.classifiers.bayes.net.estimate.SimpleEstimator) with a Laplace correction (alpha=0.5). After CPT estimation, the features were ranked based on the magnitudes of the log-odds ratios. Subsequently, derived ranks are grouped as low (0-25th percentile), moderate (25th-75th percentile), and high (>75th percentile) based on the interquartile range. Missing values were handled based on the classifier type. For non-tree-based models, missing values were imputed prior to model training using the median of the training data for numerical features and the mode for categorical features. For decision tree and rule-based classifiers, missing values were handled internally by the classifier’s default method, which distributes instances with unknown values fractionally across the branches of the decision structure based on the observed training data distribution.

Results

Age, gender, vital signs, GCS, coagulation parameters, positive CT findings, smoking history, and comorbidities were compared between survival groups. The results revealed that age, heart rate, respiratory rate, fibrinogen, international normalized ratio, and PT were significantly higher in non-survivors ($p < 0.001$ for all). On the other hand, blood oxygen saturation (SpO_2), systolic blood pressure (SBP), diastolic blood pressure (DBP), GCS, and aPTT values were significantly higher in the survival group ($p < 0.001$ for all), while no significant difference was observed between

the groups in terms of gender and platelet levels ($p > 0.05$). When comorbidities were examined, the prevalence of diabetes mellitus, hypertension, hyperlipidemia, congestive heart failure, vascular disease, malignancy, chronic obstructive pulmonary disease, chronic kidney disease, and stroke was higher in the non-survival group than in the survival group (Table 1).

Age, gender, coagulation parameters, comorbidity, and CT positivity were examined using univariate logistic regression analysis. Increases in age ($p < 0.001$), D-dimer ($p < 0.001$), fibrinogen ($p < 0.001$), and PT levels ($p = 0.009$), as well as decreases

Table 1. Comparison of the age, clinical and laboratory variables among survival groups

Variables	Distribution [†]	Survival (n=551, 88.9%)	Non-survival (n=69, 11.1%)	Overall	p
Age (years)		43.3±16.2	68.3±17.9	46.1±18.2	<0.001 ^a
Hear rate (min ⁻¹)		82 (64-150)	82 (71-152)	82 (64-152)	<0.001 ^b
Respiratory rate (min ⁻¹)		16 (12-40)	16 (14-40)	16 (12-40)	<0.001 ^b
SpO ₂ (%)		97 (8-99)	83 (45-99)	96 (8-99)	<0.001 ^b
SBP (mmHg)		120 (100-180)	120 (75-155)	120 (75-180)	<0.001 ^b
DBP (mmHg)		75 (50-110)	70 (50-90)	75 (50-110)	<0.001 ^b
GCS		15 (14-15)	15 (3-15)	15 (3-15)	<0.001 ^b
D-dimer (ng/mL)		307 (80-6610)	1173 (173-9935)	326 (80-9935)	<0.001 ^b
Fibrinogen (ng/dL)		335 (45.1-1222)	499 (79-1017)	341 (45.1-1222)	<0.001 ^b
INR		1 (0.04-24.3)	1.04 (0.87-9)	1 (0.04-24.3)	<0.001 ^b
PT (sec)		11.6 (9.4-40.1)	12.4 (1.4-100)	11.6 (1.4-100)	<0.001 ^b
aPTT (sec)		28 (16-60)	26 (16-39.5)	28 (16-60)	<0.001 ^b
Platelet (K/μL)		201 (40-684)	198 (25-457)	201 (25-684)	0.418 ^b
CT lesions in favor of COVID-19		280 (50.8%)	57 (82.6%)	337 (54.4%)	<0.001 ^c
Gender	Female	253 (45.9%)	28 (40.6%)	281 (45.3%)	0.401 ^c
	Male	298 (54.1%)	41 (59.4%)	339 (54.7%)	
DM	Yes	38 (6.9%)	21 (30.43%)	59 (9.52%)	<0.001 ^c
HT	Yes	60 (10.89%)	31 (44.93%)	91 (14.68%)	<0.001 ^c
HL	Yes	13 (2.36%)	9 (13.04%)	22 (3.55%)	<0.001 ^d
Smoking	Yes	55 (9.98%)	8 (11.59%)	63 (10.16%)	0.676 ^c
CHF	Yes	10 (1.81%)	10 (14.49%)	20 (3.23%)	<0.001 ^d
Vascular disease	Yes	19 (3.45%)	16 (23.19%)	35 (5.65%)	<0.001 ^d
Malignancy	Yes	9 (1.63%)	9 (13.04%)	18 (2.9%)	<0.001 ^d
COPD	Yes	12 (2.18%)	18 (26.09%)	30 (4.84%)	<0.001 ^d
CKD	Yes	10 (1.81%)	12 (17.39%)	22 (3.55%)	<0.001 ^d
Stroke	Yes	2 (0.36%)	3 (4.35%)	5 (0.81%)	0.011 ^d
Hepatic disease	Yes	4 (0.73%)	0 (0%)	4 (0.65%)	>0.99 ^d
Autoimmune disease	Yes	3 (0.54%)	2 (2.9%)	5 (0.81%)	0.098 ^d

[†]: Data are expressed as interquartile ranges (median, minimum-maximum) or mean ± SD, ‡: Ground glass opacities, consolidations, paving patterns, ^a: Independent t-test, ^b: Mann-Whitney U test, ^c: Pearson chi-square test, ^d: Fisher's exact test SD: Standard deviation, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, GCS: Glasgow coma score, INR: International normalised ratio, PT: Prothrombin time, aPTT: Activated partial thromboplastin time, CKD: Chronic kidney disease, DM: Diabetes mellitus, HT: Hypertension, HL: Hyperlipidemia, CHF: Congestive heart failure, COPD: Chronic obstructive pulmonary disease, CT: Computerized tomography, COVID-19: Coronavirus disease 2019, SpO₂: Blood oxygen saturation

in aPTT levels ($p < 0.001$), were associated with higher mortality. Additionally, comorbidities were dichotomized into binary groups, and patients with comorbidities had an 8.61-fold higher mortality risk compared to those with without comorbidity [odds ratio (OR)=8.61, $p < 0.001$]. Similarly, the mortality risk increased 4.6-fold among patients with positive CT results (OR=4.6, $p < 0.001$). Significant variables underwent multivariate logistic regression analysis. In the multivariate analysis, increased levels of D-dimer and advancing age continued to be associated with higher mortality rates (Table 2). GCS was excluded from these analyses because it violated the assumptions by exhibiting high multicollinearity.

ROC analysis was performed for age and coagulation parameters (D-dimer, fibrinogen, PT, and aPTT) that showed a significant association with mortality, and their predictive values for mortality were investigated. The examination revealed that at a cut-off age of 58.8 years, the sensitivity was 78.3% and the specificity was 82.6%. Among the laboratory parameters, the highest AUC value was observed in the D-dimer at a cut-off value of ≥ 638.5 ng/mL (AUC=0.832, 75.8% sensitivity and 83.1% specificity, $p < 0.001$). This was followed by fibrinogen, PT, and aPTT. The ROC analysis data for these variables are summarized in Table 3 and Figure 2.

A systematic assessment of combinations of variables and their predictive capabilities for mortality has been statistically analyzed. In step 1, the combination of age and comorbidity demonstrated a sensitivity of 81.2% and a specificity of 80.6% (AUC=0.848, $p < 0.001$). In step 2, the inclusion of vital signs into the prior combination demonstrated a sensitivity of 80.9%

and an improved specificity of 90.6% (AUC=0.905, $p < 0.001$). At step 3, age, comorbidities, vital signs, and laboratory data, which were shown to have a significant impact on outcomes (D-dimer, fibrinogen, PT, and aPTT), are integrated. Although sensitivity remains relatively unchanged at 80.4%, specificity has risen to 92.9% (AUC=0.905, $p < 0.001$). In the final stage, the clinical data (GCS and CT results) are combined, demonstrating that no substantial change is observed compared with step 3. The maximum sensitivity for mortality was 81.2%, with a specificity of 92.9%. In line with the results, considering that vital signs significantly increase specificity, whereas other combinations do not significantly increase sensitivity or specificity, vital signs are among the parameters with the greatest impact on mortality and contribute most to specificity (Table 4).

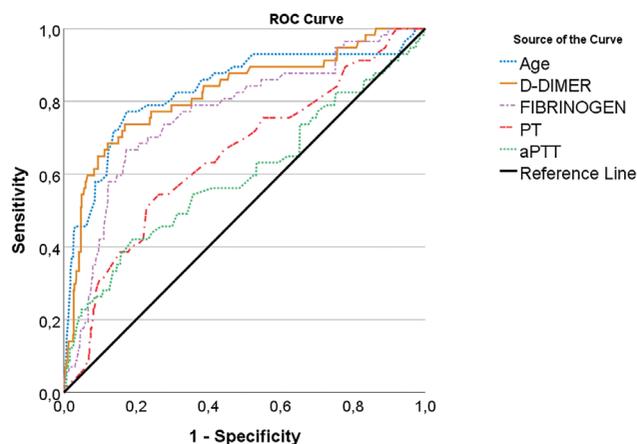


Figure 2. ROC analysis graph of variables in predicting mortality
PT: Prothrombin time, aPTT: Activated partial thromboplastin time

Table 2. Univariate and multivariate logistic regression tests for mortality

Variable	Univariate LR				Multivariate LR [†]	
	B	Nagelkerke R ²	P	OR (95% CI)	OR	P
D-dimer (ng/mL)	0.001	0.235	<0.001	1.001 (1.00-1.002)	1.00	0.029
Fibrinogen (ng/dL)	0.006	0.148	<0.001	1.006 (1.004-1.007)	1.002	0.103
INR	0.079	0.003	0.328	1.083 (0.924-1.269)	-	-
PT (sec)	0.112	0.037	0.009	1.119 (1.028-1.217)	1.061	0.221
APTT (sec)	-0.135	0.062	<0.001	0.873 (0.822-0.928)	0.944	0.111
Platelet (K/ μ L)	-0.002	0.003	0.360	0.998 (0.995-1.002)	-	-
Age (years)	0.081	0.333	<0.001	1.084 (1.065-1.103)	1054	<0.001
Gender	0.218	0.002	0.402	1.243 (0.747-2.068)	-	-
Comorbidity (overall)	2.154	0.191	<0.001	8.61 (4.78-15.54)	1.853	0.113
CT lesions in favor of COVID-19	1.525	0.086	<0.001	4.60 (2.413-8.758)	1.266	0.560

[†]: Box-Tidwell assumptions have been met and variables with high multicollinearity was excluded in the model (Nagelkerke R² of the multivariate model= 0.371). The model is built in line with the events per variable ~10 rule, COVID-19: Coronavirus disease 2019, OR: Odds ratio, CI: Confidence interval, LR: Logistic regression, INR: International normalised ratio, PT: Prothrombin time, aPTT: Activated partial thromboplastin time, CT: Computerized tomography

Table 3. ROC Analysis of parameters in favor of mortality

	AUC	95% CI		Cut-off †	Sensitivity (%)	Specificity (%)	P
		Lower limit	Upper limit				
D-dimer (ng/mL)	0.832	0.772	0.892	≥638.5	75.8	83.1	<0.001
Fibrinogen (ng/dL)	0.762	0.695	0.830	≥443.5	67.7	81.7	<0.001
PT (sec)	0.658	0.585	0.731	≥12.35	52.2	77.1	<0.001
aPTT (sec) ‡	0.631	0.550	0.712	≤24.85	43.5	84.0	<0.001
Age (years)	0.841	0.784	0.898	≥58.8	78.3	82.6	<0.001

†: Cut-off values determined based on Youden J index, ‡: Lower values are associated with positive (non-survival) results AUC: Area under the curve, CI: Confidence interval, PT: Prothrombin time, aPTT: Activated partial thromboplastin time

In our study, ML models established using 11 algorithms based on a concept suitable for our data structure were compared. The examination showed that the algorithms with the highest precision values were logistic, simple logistic, support vector machine, and multilayer perceptron. The highest overall accuracy (for combined classification of survival and non-survival groups) was achieved by the J48 algorithm. On the other hand, the best-performing algorithms in mortality prediction were BayesNet and LogitBoost. The performances of BayesNet and LogitBoost algorithms were evaluated together with F1 score, Matthew’s correlation coefficient (MCC), mean absolute error, root mean squared error, overall accuracy, Kappa (κ) and AUC values, and it was shown that the best performing algorithm was BayesNet (recall= 88.9%, AUC=88.6%, F1 score= 0.727, MCC=0.70, overall accuracy= 92.3%, mean absolute error= 0.08, κ =0.704) (Figure 3). As a result, the algorithms exhibited different predictive capabilities across different aspects, resulting in variation in recall (sensitivity), TP rate, and AUC values. The prediction features of all algorithms are summarized in Table 5.

After determining the best algorithm, the priority variables used by the algorithm and their rank scores were identified.

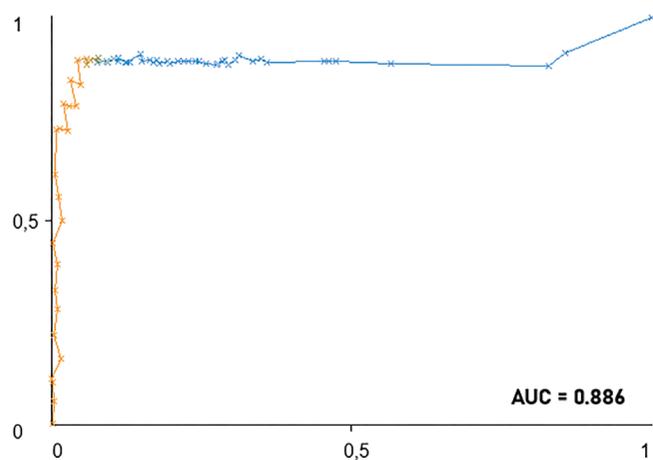


Figure 3. ROC curve of the best ML model with highest precision (BayesNet)

ML: Machine learning, AUC: Area under the curve

Thus, a ranking score was obtained for the variables used by the algorithm to determine the target class, allowing identification of variables that ranked highest in the prediction and of those with low or no effect. In the evaluation, it was understood that the most valuable parameters in mortality prediction (high-impact group) were GCS, DBP, SBP, D-dimer, SpO₂, and respiratory rate, respectively. On the other hand, the most ineffective variable, never used in prediction, was platelet level (rank score= 0). Similarly, the impact of smoking history, gender, history of hepatic disease, positive CT findings, heart rate, and history of autoimmune disease was noted to be low. The remaining attributes had moderate impacts on the prediction of the target class. The attributes and rank scores of all variables are summarized in Figure 4.

Discussion

Our study compared conventional statistical methods with ML algorithms for predicting COVID-19 mortality, showcasing the potential of ML in this complex clinical scenario. We meticulously evaluated eleven ML models, with the BayesNet algorithm demonstrating the best performance in predicting mortality (recall= 88.9%, AUC=88.6%). The J48 algorithm, although not the top performer for mortality alone, achieved the highest overall accuracy in classifying both survival and mortality groups. Importantly, the ML approach identified GCS, DBP, SBP, D-dimer, SpO₂, and respiratory rate as the most influential parameters for mortality prediction, while platelet levels were found to be least impactful (with a rank score of 0). The implementation of ML in medicine extends far beyond COVID-19 prognosis. ML algorithms offer the capability to analyze vast and complex datasets, identify non-linear relationships, and discover novel patterns that might be missed by conventional statistical methods. This leads to enhanced diagnostic accuracy, improved operational workflows, robust clinical decision support, and ultimately, better patient outcomes (3,4). As highlighted in recent reviews, ML is reshaping fields like pathology through automated image analysis, biomarker discovery, and drug development (3). The

Table 4. Stepwise evaluation of combined variables and their capabilities in predicting mortality

Combination of variables	AUC (95% CI)	p-value	Sensitivity (%) [†]	Specificity (%) [†]
Step 1: Age, comorbidity (overall)	0.848 (0.791-0.904)	<0.001	81.2%	80.6%
Step 2: Age, comorbidity (overall), vital signs ^a	0.905 (0.859-0.952)	<0.001	80.9%	90.2%
Step 3: Age, comorbidity (overall), vital signs and laboratory variables ^b	0.905 (0.851-0.958)	<0.001	80.4%	92.9%
Step 4: Age, comorbidity (overall), vital signs, laboratory variables and clinical data ^c	0.906 (0.852-0.959)	<0.001	80.4%	92.9%

^a: Heart rate, respiratory rate, SBP, DBP, SpO₂
^b: D-dimer, Fibrinogen, PT and aPTT
^c: GCS, CT lesions in favor of COVID-19
[†]: Best cut-off values is determined in accordance with Youden J index
AUC: Area under the curve, CI: Confidence interval, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, SpO₂: Blood oxygen saturation, PT: Prothrombin time, aPTT: Activated partial thromboplastin time, GCS: Glasgow coma score, CT: Computerized tomography, COVID-19: Coronavirus disease 2019

Table 5. Evaluation of ML success and prediction capabilities for each model

ML algorithms ^a	Evaluation of model success										
	Model		ML statistics - predicting mortality								
	Mean absolute error	Root mean squared error	Kappa (κ)	TP rate	FP rate	Precision	Recall	Overall accuracy [‡]	F1 score	MCC	ROC (AUC)
J48	0.094	0.236	0.723	0.778	0.036	0.737	0.778	94.2	0.757	0.724	0.842
Logistic	0.074	0.224	0.586	0.444	0.0	1.00	0.444	93.6	0.615	0.644	0.868
Simple logistic	0.08	0.213	0.530	0.389	0.0	1.00	0.389	92.9	0.560	0.60	0.880
Random forest ^b	0.186	0.233	0.736	0.722	0.022	0.813	0.722	94.8	0.765	0.737	0.861
BayesNet [†]	0.08	0.259	0.704	0.889	0.073	0.615	0.889	92.3	0.727	0.700	0.886
SMO	0.07	0.266	0.529	0.389	0.0	1.00	0.389	92.9	0.560	0.600	0.694
MultilayerPerceptron	0.08	0.267	0.443	0.333	0.007	0.857	0.333	91.6	0.480	0.503	0.884
Random tree	0.297	0.545	0.206	0.667	0.292	0.231	0.667	70.3	0.343	0.254	0.687
Decision stump	0.133	0.247	0.631	0.833	0.08	0.577	0.833	91.0	0.682	0.646	0.877
Rep tree	0.11	0.237	0.652	0.611	0.022	0.786	0.611	93.6	0.688	0.659	0.878
PART decision	0.156	0.364	0.423	0.722	0.146	0.394	0.722	83.9	0.510	0.451	0.797
LogitBoost	0.296	0.389	0.577	0.889	0.117	0.50	0.889	88.4	0.640	0.611	0.891

^a: Training and test split: 75.0%, 25.0%
[‡]: In predicting both groups (mortality and survival)
[†]: The best-performing algorithm for predicting mortality classified 16 patients (out of 18)
^b: The best-performing for overall prediction correctly classified 147 patients (out of 155)
TP: True positive, FP: False positive, MCC: Matthew's correlation coefficient, AUC: Area under the curve, SMO: Support vector machine

core strength of ML lies in its ability to learn from data, making it a powerful tool for personalized medicine, risk stratification, and early disease detection (5).

Our findings indicate that, while conventional methods, particularly stepwise logistic regression incorporating age, comorbidities, vital signs, and key laboratory data (including coagulation factors), achieved good specificity (92.9%) and sensitivity (80.4%), the BayesNet ML model achieved higher recall (sensitivity) for mortality prediction (88.9%) and a comparable AUC (88.6%). This observation is echoed in the wider literature, where

ML models often demonstrate enhanced predictive performance compared to conventional statistical models, although this is not universally the case and is context dependent (6). For instance, a systematic review and meta-analysis comparing ML with logistic regression for predicting outcomes after percutaneous coronary intervention found that ML models generally resulted in higher c-statistics, though the overall performance could be comparable, and many studies suffered from a high risk of bias (7). Another meta-analysis on predicting all-cause mortality in acute coronary syndrome patients showed that best-performing ML models had

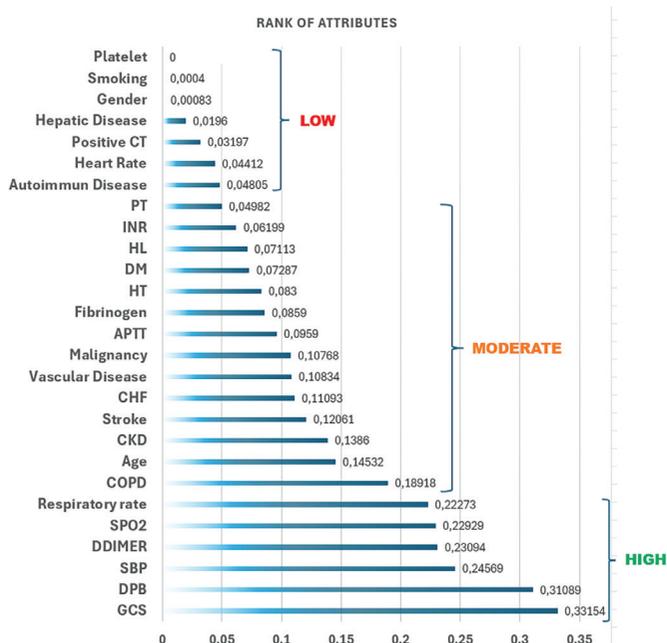


Figure 4. The worth of an attribute (feature of importance) with respect to mortality was evaluated. Ranks are statistically divided into quartiles: the top 25% is defined as “high,” the middle 25-75% as “moderate,” and the bottom 25% as “low”

CT: Computed tomography, PT: Prothrombin time, INR: International normalized ratio, HL: Hyperlipidemia, DM: Diabetes mellitus, HT: Hypertension, APTT: Activated partial thromboplastin time, CHF: Congestive heart failure, CKD: Chronic kidney disease, COPD: Chronic obstructive pulmonary disease, SpO₂: Blood oxygen saturation, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, GCS: Glasgow coma scale

a superior c-statistic (0.88) compared to conventional methods (0.82) (8). In a direct comparison between BayesNet and logistic regression, research indicates that BayesNet can be equally efficient (9) and, in certain cases, superior (10,11), particularly in managing complex variable interactions and integrating prior knowledge. The strength of conventional statistics, as used in this study, lies in its interpretability and its ability to readily analyze controlled data, though it relies on certain assumptions. ML models, while potentially more powerful in discerning complex patterns, can sometimes be perceived as “black boxes” making interpretation more challenging, a concern noted in the medical community (6). Using programming and statistical methods, this study addressed this issue by identifying the rank scores and feature importance used by the best-performing ML algorithm and by enhancing its interpretability.

This study demonstrates the prognostic significance of coagulation parameters, particularly D-dimer, for COVID-19 mortality, consistent with extensive external research. Furthermore, its comparative analysis provides valuable insights

into the superior predictive performance of ML algorithms (e.g., BayesNet) compared with conventional statistical models (logistic regression) in this setting. The ability of ML to identify key predictive variables and handle complex data interactions positions it as a promising tool for improving risk stratification and guiding clinical decision-making in COVID-19 and other complex viral diseases. Future research should focus on external validation of these ML models in diverse patient cohorts to ensure generalizability. Continued efforts to enhance the interpretability of ML models and to integrate them seamlessly into clinical workflows will be crucial to realizing their full potential in advancing patient care. The combination of robust clinical understanding, sound statistical principles, and advanced computational techniques, as exemplified in this work, paves the way for more accurate and timely medical predictions.

While this study utilized a COVID-19 dataset due to its clinical richness and the heterogeneous nature of the disease, the findings offer broader insights into methodology selection for predicting outcomes in similar clinical contexts. Our study not only emphasizes the need for other similar conceptual studies, but also serves as an ideal guide for the research and development of ML applications in the field of medicine.

Study Limitations

A limitation of our study is that the sample size of the output class (non-survival group) was inadequate to fit multivariate logistic regression models that included all variables. Therefore, we had to limit the events per variable rule to approximately 10, which required us to exclude some variables from the model. Additionally, unlike ML models, the statistical methods we employed required adherence to certain assumptions (e.g., Box-Tidwell test and multicollinearity). Consequently, we excluded variables that did not satisfy these assumptions. Another limitation of our study is the lack of external validation to assess the efficacy of the ML algorithms. The Nagelkerke R² value of 0.371 in the multivariate logistic regression model indicates that the current variables account for only 37.1% of the variance in mortality. This suggests the presence of many additional parameters that could be associated with mortality, highlighting the complex nature of COVID-19. These findings indicate that further studies to identify additional variables, with confounders defined may be necessary. Further assessment of ML algorithms and their comparisons with conventional statistical methods are required.

Conclusion

Conventional statistical methods require that certain assumptions be met and have limitations. With more algorithmic structures

and their ability to process complex datasets, ML algorithms can achieve higher sensitivity in mortality prediction. For complex, multifactorial diseases, ML methods are emerging as a promising approach.

Ethics

Ethics Committee Approval: The study protocol was designed in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Selçuk University Rectorate Local Ethics Committee of a university hospital (approval no: 2025/218, date: 08.04.2025).

Informed Consent: Informed consent was waived due to the retrospective design.

Footnotes

Authorship Contributions

Concept: F.İ., A.A., Design: F.İ., Data Collection or Processing: F.İ., A.A., Analysis or Interpretation: F.İ., Literature Search: F.İ., Writing: F.İ.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Nørskov AK, Lange T, Nielsen EE, Gluud C, Winkel P, Beyersmann J, et al. Assessment of assumptions of statistical analysis methods in randomised clinical trials: the what and how. *BMJ Evid Based Med.* 2021;26:121-6. Epub 2020 Jan 27.
2. Ahsan MM, Luna SA, Siddique Z. Machine-learning-based disease diagnosis: a comprehensive review. *Healthcare (Basel).* 2022;10:541.
3. Hanna MG, Pantanowitz L, Dash R, Harrison JH, Deebajah M, Pantanowitz J, et al. Future of artificial intelligence-machine learning trends in pathology and medicine. *Mod Pathol.* 2025;38:100705. Epub 2025 Jan 4.
4. Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920-30.
5. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. *Med.* 2021;2:642-65. Epub 2021 Apr 30.
6. Mann J, Lyons M, O'Rourke J, Davies S. Machine learning or traditional statistical methods for predictive modelling in perioperative medicine: a narrative review. *J Clin Anesth.* 2025;102:111782. Epub 2025 Feb 19.
7. Nayebirad S, Hassanzadeh A, Vahdani AM, Mohamadi A, Forghani S, Shafiee A, et al. Comparison of machine learning models with conventional statistical methods for prediction of percutaneous coronary intervention outcomes: a systematic review and meta-analysis. *BMC Cardiovasc Disord.* 2025;25:310.
8. Gupta AK, Mustafiz C, Mutahar D, Zaka A, Parvez R, Mridha N, et al. Machine learning vs traditional approaches to predict all-cause mortality for acute coronary syndrome: a systematic review and meta-analysis. *Can J Cardiol.* 2025;41:1564-83. Epub 2025 Feb 17.
9. Ducher M, Kalbacher E, Combarrous F, Finaz de Vilaine J, McGregor B, Fouque D, et al. Comparison of a Bayesian network with a logistic regression model to forecast IgA nephropathy. *Biomed Res Int.* 2013;2013:686150. Epub 2013 Nov 17.
10. Yuan B, Wang C, Fan Z, Liu C, Fang L, Ma L, et al. A Bayesian network-based approach for identifying risk factors and predicting ischemic stroke in infective endocarditis patients. *Front Cardiovasc Med.* 2024;10:1294229.
11. Ma L, Cai B, Qiao ML, Fan ZX, Fang LB, Wang CB, et al. Risk factors assessment and a Bayesian network model for predicting ischemic stroke in patients with cardiac myxoma. *Front Cardiovasc Med.* 2023;10:1128022.